Why is behavioral game theory a game for economists? The concept of beliefs in equilibrium*

Michiru Nagatsu Chiara Lisciandra†

May 7, 2021

Abstract

The interdisciplinary exchange between economists and psychologists has so far been more active and fruitful in the modifications of Expected Utility Theory than in those of Game Theory. We argue that this asymmetry may be explained by economists' specific way of doing equilibrium analysis of aggregate-level outcomes in their practice, and by psychologists' reluctance to fully engage with such practice. We focus on the notion of belief that is embedded in economists' practice of equilibrium analysis, more specifically Nash equilibrium, and argue that its difference from the psychological counterpart is one of the factors that makes interdisciplinary exchange in behavioral game theory more difficult.

1 Introduction

One of the most influential texts published in the behavioral and social sciences in the first half of the twentieth century was, according to many, Von Neumann and Morgenstern's (vNM) *Theory of Games and Economic Behavior* (1944). Not only did the book lay the foundation of game theory, which has become the essential research tool in contemporary economics, it also influenced several other disciplines beyond economics, from political science to linguistics and biology.

Two of the most important contributions of Theory of Games and Economic Behavior to economics are the axiomatic derivation of expected utility (in the second edition of 1947) and the minimax solution to zero-sum games. John Nash generalized vNM's existence proof of equilibrium in non-zero-sum games in 1951, thereby

*To appear in Egashira, Taishido, Hands and Mäki (eds.) (2018) *A Genealogy of Self-interest in Economics* (Springer). We thank Wade Hands, one of the editors, for valuable comments.

†Both authors equally contributed to research and writing of the article.

providing game theory with a foundational solution concept. Leonard Savage completed the axiomatization of expected utility theory with subjective probabilities in 1954, thereby putting the standard formal approach to decision-making under risk.

During the years following these publications the axiomatic treatment of subjective expected utility theory has been modified substantially through interactions between psychologists and economists, whereas the modification of Nash equilibrium has been left to economists for the most part. What explains this asymmetry? Our aim in this chapter is to offer some methodological reasons for the differences in the scholarly reception of the two main components of *Theory of Games and Economic Behavior*.

Among the first reactions to the formalization of decision and game theory were a number of thought experiments that sparked a debate over their normative and/or descriptive status. Classic examples include Allais's paradox, Ellsberg's paradox, and Schelling's focal points. Soon after that, an increasing number of laboratory experiments were conducted to systematically test the predictive accuracy of subjective expected utility theory, i.e., the individual-decision-theory side of the book. Several cognitive psychologists, in particular Lichtenstein and Slovic (1971), and Kahneman and Tversky (1971; 1979) reported the first critical results. Economists took these results of laboratory and field experiments seriously, and started to design experiments to test the robustness of the psychologists' findings (see e.g., Grether and Plott (1979)). Hence, an interdisciplinary community of economists and psychologists proved pivotal to the rise of behavioral decision theory during the 1970s and 1980s.

On the other hand, the strategic side of the book has not prompted the same pattern of interdisciplinary exchange between economics and psychology. Although it has influenced many other disciplines in different ways, game theory has not been modified by psychology, as decision theory was. As a consequence, exchanges between economics and psychology on the basis of the common use of game theory remain limited to this day.

This is surprising, because in many ways game theory and expected utility theory share major features—the mathematical proofs of existence and uniqueness are foundational for both. In fact, expected utility theory was originally devised by von Neumann and Morgenstern to solve *games* with strategic uncertainty.

Our aim in this chapter is to consider the features of game-theoretic models that prevented them from being fully informed by psychological research, relative to decision-theoretic models. We argue that these features concern the belief concept embedded in the practice of equilibrium analysis that is specific to economics. In particular, we argue that the asymmetry in the reception of *Theory of Games and Economic Behavior* is attributable in part to the notion of belief implied in the Nash equilibrium.

We proceed as follows. Section 2 discusses and specifies the asymmetries in more detail. Section 3 presents two emblematic episodes that illustrate how the concept of equilibrium does not pertain to the explanatory toolbox of psychologists, whereas it is essential for economists. Section 4 identifies beliefs as a key construct that appears in both expected utility theory and game theory, but is understood differently: we argue that this mismatch made knowledge exchange between economists and psychologists more difficult. Section 5 explicitly analyzes expected utility theory and game theory as *boundary objects*—means of interdisciplinary exchange—and clarifies their asymmetries by drawing on the previous sections. Section 6 briefly concludes the chapter.

2 Asymmetries

This section concerns the asymmetries between behavioral game theory and behavioral decision theory, with respect to their development as research programs and their impact on the scientific community. We focus first on the debate concerning the normative/descriptive status of GT, arguing that even if it evolved differently than in the case of EUT, the interpretation of GT as a descriptive theory still has a legitimate theoretical status and practical relevance, as the development of behavioral game theory shows. This point rules out an alternative hypothesis explaining the asymmetric developments of behavioral decision theory and behavioral game theory, namely that EUT allows straightforward empirical interpretation whereas GT is a normative or prescriptive theory, whose empirical status is complex and contested.

We then substantiate the asymmetry between the two fields with reference to the different role of psychologists in the genesis of experimental work. In doing so, we reject another possible explanation of the asymmetry suggesting that psychologists did not find a fertile ground for collaboration in GT because the experimental results could all be accommodated within a framework of pure rationality. Our ultimate goal in this section is to demonstrate the need for a fundamental explanation of the difference between the two fields, and we argue that this is to be found in the different concepts of beliefs they adopt.

2.1 Indications of asymmetry and alternative explanations

Some initial empirical support for the asymmetry we identify is provided in a recent paper by Doehne and Herfeld (2018) in which co-citation network analysis is used to study the diffusion of scientific innovations introduced by von Neumann and Morgenstern (1944). The authors show how the spread of TGEB was mediated by scientific publications acting as "translators" to other fields, defining translators

as publications that were pivotal in showing the relevance and potential of the innovations for a particular domain, sometimes creating a new field of inquiry as a result.

Doehne and Herfeld's analysis clearly shows that the diffusion of rational choice theory to psychological fields was mediated by psychologists as translators. Interestingly, however, all the fields these translators created or impacted concern expected utility theory (behavioral decision theory, statistical decision theory, and mathematical psychology), but not game theory. There is, in fact, no psychological translator in the domains of cooperative game theory, non-cooperative game theory, or theories of conflict and cooperation.

This asymmetry in how decision theory and game theory have spread is puzzling if one considers that a similar pattern seems to characterize the way in which researchers initially received EUT and GT: both were also interpreted as descriptive theories yielding predictions that could be tested experimentally (see below). Indeed, in the opinion of many, the empirical results were problematic for both. Specifically, preference reversals, framing effects, fair divisions in the ultimatum game, and cooperation in the public goods game were challenging outcomes for the status of both EUT and GT as predictive and empirical theories.

A common reaction at this point is to say that GT has not promoted interaction between economists and psychologists, because GT should not be taken as a descriptive theory. However, such an explanation downplays some significant aspects of the research program GT has prompted. For one thing, the descriptive side of GT is flourishing, as the development of behavioral game theory indeed shows. If the reason why psychologists have not collaborated with game theorists is that GT is a normative theory, then we should not observe the development of behavioral game theory, either. On the other hand, behavioral game theory is a growing research area that has developed without a decisive contribution from psychology. Moreover, many of the reasons why a normative reading would be more legitimate than its descriptive counterpart are discussed in the literature without leading to the conclusion that the normative interpretation of GT is the only legitimate one (on this point, see (Guala, 2006)).

Another possible reaction is that the more complicated or contested empirical status of GT compared to EUT explains the asymmetry. The debate concerning the interpretation of empirical results in GT has indeed been more complex and contentious than that concerning EUT. One of the main reasons for this is that GT has a richer structure: it involves more than one individual decision maker, and therefore requires specification of the number of players, their interrelated strategies, related payoffs, and the information they have. This makes the problem of underdetermination more severe: when the empirical results diverge from theoretically derived predictions, there are more potential culprits.

Although empirical tests of EUT also have to deal with the problem of under-determination (whether it is risk, loss, or regret that people avoid; whether it is beliefs rather than preferences), at least experimenters do not have to consider the options that derive from the richer structure of GT, as described above. In particular, they do not have to consider the utility involving beliefs about others in that decision problems concerning lotteries do not involve other players. In short, tests of GT suffer more severely from under-determination issues than tests of EUT do.

The complex structure of GT further challenges its empirical status, as shown in the problem concerning the refinement of payoffs, or of what to include in *self-interest*. By way of an illustration, let us suppose that the participants in a laboratory one-shot Prisoner's Dilemma game decide to cooperate, contrary to the theoretical prediction of mutual defection. In this case, the payoffs could be refined to include the broader self-interest of the participants, such as preferences that satisfy the other's preferences or perceived expectations, so that mutual cooperation is still explained as a result of rational play. The underlying justification is that this enables the experimenter to incorporate non-monetary outcomes that people care about. The worry, however, is that this strategy might make it impossible to test the theoretical predictions of rational play as opposed to selfish or narrowly-self-interested play. Not only would this make the theory unfalsifiable, it would also imply that the entire research program on behavioral game theory is misconceived, if not flawed.

However, concerns that game theory is empirically vacuous can be addressed (see e.g., Guala, 2006). The refinement of people's utility function can be empirically justified to the extent that it captures more or less stable *patterns* of choices. In other words, it is possible to impose some empirical discipline on *ad hoc* postulations that any observed behavior is self-interested or maximizes utility in one way or another. In this respect, the rationales underlying experimental work and modifications in GT and in EUT are no different. In both cases, experimental analysis is supposed to test the choice models of individuals and, if necessary, provide indications of how to formulate more accurate descriptive models. Models of social preferences in behavioral game theory have been empirically evaluated in exactly this way, for example (Camerer, 2003)..

Our first conclusion is thus that the asymmetric developments of behavioral decision theory and behavioral game theory are attributable neither to the distinctly normative or prescriptive character of game theory's nor to its complex or contested empirical character compared to expected utility theory. Although the debates on EUT and GT proceeded independently, the goals of experimental work were similarly conceived. The main point in both cases is to test the predictions and, in case of divergence, suggest how to modify the theory so as to accommodate them in a

systematic way.

2.2 The genesis of experimental work on game theory and decision theory, and how the roles of psychologists compare

Let us now turn to the genesis of experimental work. There were differences in the way in which experiments were introduced in EUT and GT—and in the responses they provoked. First, the body of experimental work on EUT that led to the development of behavioral economics was, in the main, instigated by psychologists. Economists took up the challenges raised in the empirical work almost immediately, and fruitful exchanges took place between the two disciplines.[1] The most influential theory to date remains prospect theory, which was developed by two psychologists.

In the case of GT, however, psychologists did not have as influential a role as they had in promoting experiments on EUT and in the following theoretical developments. Two mathematicians, Merrill Flood and Melvin Dresher, conducted the first experiments on the prisoner's dilemma in 1950. Another milestone paradigm in experimental game theory, the ultimatum game, was developed by three economists—Werner Güth, Rolf Schmitterberger and Bernd Schwarze—in the context of testing Ariel Rubinstein's model of bargaining.[2] After this, the ultimatum game provided a testbed for Fehr and Schmidt's iniquity-aversion model in *A Theory of Fairness, Competition, and Cooperation* (1999), which has become extremely influential among behavioral and experimental economists.

A different story holds for social dilemma and public good games, which indeed have attracted social psychologists, sociologists and experimental economists, generating a lively multidisciplinary field (for a review, see Ledyard, 1995). Here in fact, is a pattern that closely resembles what happened in the experimental testing of EUT.

The interest of the economists—in particular of Mark Isaac and James Walker—in public good games was initially sparked by the experimental findings of a group of psychologists, under the guidance of Robin Dawes, and a group of sociologists including Gerald Marwell and Ruth Ames. Both groups were working on the so-called social dilemmas, which are a special version of public goods games (Dawes et al., 1977).[3]

[1] See Lisciandra (2018) for an analysis of experiments on social preferences.

[2] Daniel Kahneman was very "crestfallen" (Kahneman, 2014) when he got to know that he and his economists colleagues had been scooped by Güth et al. Kahneman published the scooped study as Kahneman et al. (1986)

[3] More generally, research on non-cooperative game theory—most importantly on Prisoner's Dilemma and social dilemma—formed the field of conflict resolution. The field is problem-oriented and less theoretically integrated; that is, the field involves researchers from multiple disciplines

The surprising observation among economists was that social dilemmas seemed not to suffer from the problem of free-riding behavior, as game theory predicts. Tying to prove them wrong, Isaac and Walker ended up with mixed results: their work did not reverse previous findings, but it did not confirm them, either. One major innovation that these economists brought to the table was the repeated game design, which turned out to influence individual contributions in experimental settings in interesting ways. The experiments indicated that, at least under certain conditions, free-riding behavior emerges with repetition (Kim and Walker, 1984; Isaac et al., 1985) in a way that is consistent with the theoretical expectations (coupled with an assumption that people initially make mistakes and eventually learn to play rationally).

Nevertheless, the results from public goods games were less conclusive than was desirable (see Ledyard, 1995), and the overall picture that emerged from the empirical work was rather challenging for game theory. Among the main puzzling aspects, there was the issue of how to conceptualize fairness considerations or focal points in non-cooperative games, and how to explain the solutions to coordination games that have multiple Nash Equilibria, such as the Hi-Lo game.[4]

In a similar way as with EUT, these results seem to show that a purely game-theoretic approach that focuses exclusively on the rationality criteria is not well equipped to explain the phenomena observed in experiments. It may be that a more careful look at the process of beliefs formation could help to solve the foundational difficulties facing game theory (see Colman, 2003).

The question thus remains: why have psychologists not engaged more critically with rational models of game theory in a way that would lead to theoretical modifications? After all, beliefs and desires lend themselves to psychological analysis, as the case of EUT shows.[5] Moreover, given that beliefs are at least as crucial in GT as in EUT, it is even more surprising not to observe the same pattern of exchange in the former. As we have argued above, the normative versus the descriptive interpretations cannot explain the divide between EUT and GT, hence the need for a different, more fundamental explanation of the asymmetry.

Below we will argue that one of the reasons for this outcome is that the very notion of beliefs is interpreted and used differently in the two theories. In the case of GT, desires are captured in the payoff of outcomes corresponding to strategy profiles, or different combinations of players' actions. When players choose actions

such as economics, psychology, political science, law and management (see Mnookin et al., 1995), but there has not been significant theoretical integration of insights from these disciplines.

[4]The Hi-Lo game is a game with two Nash equilibria, one of which has higher payoffs for both players than the other. Game theory, however, treats the two solutions as equally attractive, because strictly speaking they are both Nash equilibria. See Bacharach et al. (2006)

[5]It has indeed been argued that the "journey" from EUT to psychology and back was facilitated by the familiarity of psychologists with such concepts (Nagatsu and Małecka, 2019).

such that their payoffs are maximized, the game is said to be "solved" because the system is in an equilibrium state—no unilateral deviation from that state yields benefit: the beliefs of each players about the other player's beliefs and actions are indeed correct. However, in many cases the analysis is silent about the way in which these beliefs have been reached: it finds a resting point, but it leaves unanswered the question of whether and how such a state can be reached by real players. Equilibrium analysis thus understood is a very familiar exercise to economists, but it is not equally prominent in the scientific practice of psychologists. If anything, psychologists are interested in *how* the equilibrium is reached, rather than in its theoretical foundations in strategic interactions *per se*. As we show in the following sections, this difference might have determined the divide between the two groups, and it could explain the asymmetries we have identified above.

3 Equilibrium analysis: two illustrations

As we briefly noted at the end of the previous section, a crucial difference between the belief concept in GT and its counterpart in EUT is that the former is deeply connected to the equilibrium analysis of interactive play, whereas the latter is not. To understand this difference, we need first to characterize equilibrium analysis in economics, and then to illustrate how it figures in explanatory practices.

Equilibrium analysis, in its most abstract sense, is an approach to studying a given system's emergent properties as a result of interactions between its components. A market consisting of consumers and producers, and the markets comprising individual markets, have been the the main systems of interest in economics.[6] Game theory has extended economic analysis to any systems involving two or more interactive agents, and has shifted the focus from dynamic optimization to the mutual consistency of the system's components (Giocoli, 2003). As this shift implies, equilibrium analysis as applied by economists has changed in character during the history of economics.[7] Nevertheless, it is possible to characterize economic equilibrium analysis as an approach to studying the behavior of a system (or a set of systems) primarily in relation to its well-defined stable states. Our claim is not that only economists engage in this mode of analysis (in fact psychologists engage

[6]Studying optimization of consumers and firms as a result of profit and utility maximization under budget constraints is a type of equilibrium analysis, in which the systems in question are individual agents. In this chapter, however, we focus on the equilibrium analysis involving multiple agents.

[7]Hands (2010) argues that, independently from game theory, in the mid 20th century a significant change in the character of equlibrium analysis has occurred in consumer choice theory and general equilibrium theory in economics: the shift from finding a stable point toward which a system moves through a certain path, to finding a rest point in an dynamically related system of differential or difference equations.

similarly in the study of bounded rationality), but that economists' way of doing so in game theory has created a specific disciplinary barrier in changing the nature of the belief concept in a non-transparent way.

To illustrate this point, we present two episodes contrasting responses of economists and psychologists to equilibrium analysis. The first one comes from game theory and the second one from experimental economics. We show how psychologists are reluctant to engage with that kind of explanations, then we compare the logic behind equilibrium-based explanations in game theory with optimization in decision theory. Let us start with the theoretical case.

3.1 The by-stander effect

The first case concerns the so-called by-stander effect in social psychology. In a popular introductory textbook on game theory, Martin Osborne (2004) builds a model of "reporting a crime," in which a group of homogeneous people decide whether or not to report a crime they have observed. The model serves as an illustration of how to use mixed-strategy equilibrium to solve coordination games with conflict (Section 4.8 in Osborne, 2004).

The game is played by n players, whose action set is {Call, Don't call}, and whose preference ordering over three outcomes is as follows: someone else calls $\succ$ she calls $\succ$ no one calls; mapping to expected utilities $v > v - c > 0$ respectively, where v is the value she attaches to the crime being reported, and c is the cost she incurs to call herself, and $v > c > 0$. The model is intended to explain the brutal murder of Catherine ("Kitty") Genovese over a period of half an hour in New York City in 1964, discussed in the textbook.

The puzzle to be explained is why none of the 38 people who witnessed the incident reported the crime to the police. Osborne's answer is that, other things being equal, the mixed-strategy Nash equilibrium of the game implies that the higher the number of people who witness the incident, the more likely it is that no one will report it. Specifically, it is an implication of the equilibrium condition that each player must be indifferent between calling and not calling in equilibrium:

$$v - c = v * \Pr\{\text{at least one other person calls}\} + 0 * \Pr\{\text{no one else calls}\}$$

or

$$v - c = v * (1 - \Pr\{\text{no one else calls}\})$$

or

$$\Pr\{\text{no one else calls}\} = c/v$$

Let us denote the probability that each person calls as p. The probability that no one else calls is the probability that every one of the $n - 1$ people does not call,

i.e., $(1-p)^{n-1}$. The equilibrium condition is $(1-p)^{n-1} = c/v$, or

$$p = 1 - (c/v)^{1/(n-1)}$$

This is the unique, symmetric mixed-strategy equilibrium of the game, in which the probability of each person calling is p $(1 > p > 0)$. Given that this probability decreases as n increases, it is clear that the probability of each person's reporting decreases as the group becomes larger. The more subtle point—that the probability that *no one* will report increases as the group becomes larger—is shown by focusing on any player i.

$$\Pr\{\text{no one calls}\} = \Pr\{i \text{ does not call}\} \times \Pr\{\text{no one else calls}\}$$

Recall that $\Pr\{\text{no one else calls}\} = c/v$, which is independent of the group size n. Because $\Pr\{i \text{ does not call}\}$ increases as n increases (this is just $1-p$ for i), one could conclude that the probability that no one calls also increases as n increases. The crime was unreported in the Genovese case *because of*—not despite—the large number of witnesses. In other words, it is better to have fewer people around if we hope to be rescued!

Osborne contrasts his explanation to three others offered by social psychologists to explain similar experimental findings about by-stander effects. The first one concerns the diffusion of responsibility—the larger the group size, the smaller is the psychological cost of not helping; the second is called audience inhibition—the larger the group, the greater the potential embarrassment of a helper if the help turns out to be inappropriate; the third one is about social influence—the larger the group size, the more likely it is that witnesses will infer from the inaction of the others that help is, in fact, not appropriate. Osborne subsumes these explanations in his model as the group size (n) either raising the expected cost (c) or reducing the expected benefit (v) of helping.[8] He then points out that the implication of a mixed Nash equilibrium—that the larger the group size the less likely it is that at least one person will report a crime—holds even if the values of v and c are *independent* of the group size n or, equivalently, even if the three psychological effects of group size are absent.

Osborne's point, therefore, is not that these psychological explanations are wrong or redundant—some or all of them may well be contributing factors—but rather that they all miss the crucial notion of an *equilibrium*. Osborne thus concludes:

> Whether any given person intervenes depends on the probability she assigns to some other person's intervening. In an equilibrium each person must be indifferent between intervening and not intervening, and as

[8] In fact, the third explanation can be cast in a game theoretic model of pluralistic ignorance, but we will follow Osborne's presentation here.

> we have seen this condition leads inexorably to the conclusion that an increase in group size reduces the probability that at least one person intervenes. (pp. 133–134)

For the current purposes, it is not important whether Osborne's is the correct explanation of the by-stander effect in general, or of the Genovese case in particular. Nor is it important whether his account is more unifying than those offered by psychologists in some sense that philosophers of science specify. The moral of this illustration is rather that Osborne clearly contrasts equilibrium analysis that is central in economics to typical social-psychological explanations with no explicit equilibrium analysis. Note that the asymmetry does not imply that equilibrium analysis excludes any concepts of beliefs. On the contrary, Osborne states that the agent assigns a probability to an event in order to be indifferent between two actions. This point requires some mental ascription to the agent he is modeling. However, as we illustrate in the next section, these psychological concepts, beliefs in particular, are understood and used in economics very differently than in psychology. Before discussing that, we present another episode that highlights the specificity of equilibrium analysis.

3.2 Almost like magic: the N^* game

Another illustration of how equilibrium analysis is distant to psychologists comes from Nobel Laureate Daniel Kahneman, arguably the most authoritative psychologist to talk about economics. In his 2002 autobiography, he recalls the "magic" he observed in an experiment he conducted with economists Richard Thaler and James Brander. The experiment, called the N^* game, is an N-player symmetric market-entry coordination game without communication: the market is profitable for entrants, but the marginal profit from entry decreases as the number of entrants increases, and beyond a certain market capacity (denoted by N^*) profit becomes negative. In the original experiment, $N = 15$ and N^* was changed over a period of repetition within the range of $12 > N^* > 3$; the payoff to each person was \$.25 if one did not enter the market, and $\$[.25 + .5(N^* - E)]$ if one did, where E is the number of total entrants.[9] If $E = N^*$—if the number of actual entrants equals the capacity—both entrants and non-entrants receives the same \$.25 payoff. To Kahneman's great surprise, E quickly converged to N^* in a few rounds, and stayed within the range $N^* - 2 < E < N^* + 2$ in the vast majority of trials. Of course, this is the implication of a mixed-strategy Nash equilibrium: each player decides to enter with a probability such that the expected payoff from entry equals that from

[9] We used the numbers as in Kahneman (1988), which are slightly different from those presented in his Nobel authobiography.

non-entry—and is indifferent between entering and not entering in a steady state. The aggregate outcome from such individual strategies will result in $E \approx N^*$.

In Kahneman's words, "[o]bserving the regularity of behavior in these markets was a bewildering experience—to a psychologist, it looked almost like magic." (1988, p. 12) And "it took me some time to realize that the magic we were observing was an equilibrium: the pattern we saw existed because no other pattern could be sustained." (2014) Moreover, the debriefing conversations revealed that most participants' accounts of their own "winning" strategies were unfounded and had no clear connection with the equilibrium results. In other words, "[t]he equilibrium outcome (which would be generated by the optimal policies of rational players) was produced in this case by a group of excited and confused people, who simply did not know what they were doing." (Kahneman, 1988, p. 12)

Kahneman summarizes his lessons from this study as follows:

> Psychologists are trained to believe that aggregate phenomena can be explained by finding some relevant regularity in individual behavior. The N^* game provided me with first-hand experience of a clear failure of this belief. The only solid explanation of the results of the N^* game belongs to a type that is quite familiar to economists, but not to other social scientists.[...] The cognitive psychologist discovers that he has essentially nothing of interest to contribute, and that his bag of intellectual tools lack the powerful instrument of equilibrium explanations. (Kahneman, 1988, pp. 12–13)

This is a surprisingly unguarded remark, coming from the cognitive psychologist who has so forcefully and successfully challenged EUT, or the vNM utility notion used in the derivation of the mixed-strategy Nash equilibrium. One could argue that the results are also problematic for economists because they cannot explain the gap between the equilibrium results and the (confused) self-reports of the participants of the N^* game, either. However, in that economists typically neither demand players' conscious awareness of their reasoning processes, nor value self-reports as reliable evidence, this gap seems less problematic for them. In any case, equilibrium analysis yields accurate prediction, whereas no psychological account alone fills the gap.[10]

As in the Osborne case, the fact that equilibrium analysis explains some phenomena, whereas psychological accounts do not does not imply that the former involves no psychological concepts. As we will show in the next section, the concept of beliefs is implicit in the use of the mixed-strategy Nash equilibrium.

[10]For modern discussion on this game, see Dhami (2016). Interestingly, no other behavioral economics textbooks than this advanced one discusses N^* game.

3.3 The methodological rationale behind equilibrium analysis

The two episodes described above encourage the use of equilibrium analysis in showing how successful it can be as an explanation or prediction. This is not the only reason behind the use of equilibrium-based accounts, which do not always succeed in predicting or explaining the phenomena of interest, be they market aggregates or collective behaviors. Economists therefore resort to a more general, methodological rationale. According to Herbert Gintis, for example, economists explain aggregate phenomena in terms of an equilibrium "not because it accurately reflects actual economic conditions, but rather because it is instructive to understand when it does not, and why" (Gintis, 2017, p. 251). In other words, equilibrium analysis is justified as providing an empirical benchmark and a heuristic for explaning deviations when they exist.

In fact, this methodological rationale is very similar to the rationale behind the use of *optimization* as a benchmark in behavioral decision theory, in which a clear prediction from the canonical optimization model (e.g. EUT or exponential time-discounting) is derived, then a deviation is experimentally demonstrated, and finally explanations in terms of individual psychological biases (e.g., loss aversion or present bias) are provided. It is also possible to extend the same "benchmark-deviation-biases" strategy to explaining aggregate-level phenomena. In their famous mug cup study, for example, Kahneman et al. (1990) first derived a clear prediction from market-equilibrium analysis: because mug cups are *randomly* given to half of a group of the participants, there is a 50-50 chance that the new owners will value their cups more than those who did not receive one. Therefore, half of the mug cups would be voluntarily traded (benchmark). The researchers then demonstrated that the volume of the trade was significantly less than half (deviation), which they finally explained in terms of loss aversion or anticipated regret making the cup owners more reluctant to sell relative to the willingness to pay of potential buyers (bias). The endowment effect thus constructed is an aggregate-level phenomenon, but it is explained in terms of individual bias.

The upshot of this section is that, although equilibrium analysis is a very economic way of explaining aggregate phenomena, compared to social and cognitive-psychological approaches, its underlying methodological rationale (benchmark and heuristic) is no different from that of psychologists engaged in behavioral decision research. In what sense, then, are these two illustrations of game-theoretic equilibrium analysis alien to psychologists? As we hinted in this section, we now argue that the conceptual difference between beliefs in equilibrium and beliefs in individual decision-making may have been the main barrier to productive collaboration (in a broad sense) between economists and psychologists in behavioral game theory.

4 Beliefs in equilibrium and beliefs in individual decision making

In the previous section we presented two case histories, one of a formal model (told by an economist) and the other of an experiment (told by a psychologist), both of which highlight equilibrium analysis as a distinctively economics-based explanatory style of observed aggregate patterns of human behavior. However, this does not mean that equilibrium analysis is void of psychological constructs. On the contrary, finding an equilibrium solution to a given game necessitates conceptualization of beliefs as a theoretical construct, as we will show. Our point is rather that such a belief concept is distinct from its psychological counterpart that features in EUT. our aim in this section is to make this conceptual difference explicit. First we explain how the belief concept is defined in game theory, and how it is implied in the way equilibrium analysis is conducted, then we contrast it to its counterpart in EUT.

Hal Varian (2010), in his popular intermediate-level microeconomics textbook, informally defines a Nash equilibrium in a two-player strategic form game as follows:

> a pair of strategies is a **Nash equilibrium** if A's choice is optimal, given B's choice, *and* B's choice is optimal given A's choice. (p. 524)

Here is Osborne's (2004) slightly more technical and precise definition in the introductory textbook on game theory discussed above:

> *A Nash equilibrium* is an action profile a^* with the property that no player i can do better by choosing an action different from a_i^*, given that every other player j adheres to a_j^*. (p. 22)

Although beliefs do not feature in these definitions, Varian (2010) makes a significant observation immediately following his definition quoted above.

> Remember that neither person knows what the other person will do when he has to make his own choice of strategy. But each person may have some expectation about what the other person's choice will be. A Nash equilibrium can be interpreted as a pair of expectations about each person's choice such that, when the other person's choice is revealed, neither individual wants to change his behavior. (pp. 524–5)

In other words, players do not know what the other person will do, but in equilibrium they act *as if* they did: their beliefs about others' actions are coordinated in the sense that they are the same and true in equilibrium. In fact, coordinated beliefs are the second component in the notion of Nash equilibrium, as Osborne (2004, p. 20) explicitly mentions, and they are made even more explicit in Varian's (1992) advanced microeconomics textbook (p. 265):

A Nash equilibrium consists of probability beliefs (π_r, π_c) over strategies, and probability of choosing strategies (p_r, p_c), such that:

1. the beliefs are correct: $p_r = \pi_r$ and $p_c = \pi_c$ for all r and c; and,
2. each player is choosing (p_r) and (p_c) so as to maximize his expected utility given his beliefs.

Varian refers to two players, Row and Column; p_r denotes the probability of Row playing r; and π_c denotes Row's subjective probability distribution over Column's choices, i.e. Row's beliefs about Column's behavior, and similarly for Column. Of note here is that this definition requires each player's subjective beliefs about others' choices to coincide with their actual choices.

Varian warns the advanced reader that a more conventional definition of a Nash equilibrium—such as his in the intermediate textbook and Osborne's quoted above—is misleading "since the distinction between the beliefs of the agents and the actions of the agents is blurred" (1992, p. 266). What, then, is the nature of these beliefs thus distinguished from the actions of agents? Do agents really "know" other agents' actions in some philosophically or psychologically well-founded sense?

These questions concern the interpretation of a Nash equilibrium and are not part of any formal definition. In fact, during the 1980s there was a debate between psychologists and game theorists concerning the exact interpretation of beliefs in game theory. In brief, psychologists such as Kadane and Larkey (1982) criticized the assumption that players' beliefs about each others' beliefs and behavior in equilibrium were necessarily true, whereas game theorists such as Harsanyi defended this interpretation of beliefs as necessary theoretical apparatus to derive a solution to any given game (see Morris, 1995; Grüne-Yanoff and Lehtinen, 2012, for a detailed discussion of the reasons why economists adopt the common prior assumption).

In contrast to this philosophical and theoretical literature on the nature and epistemological foundations of beliefs in game theory, the pronouncements of many practicing economists on these issues are neither explicit nor eloquent. In fact, there is some indication that economists are not primarily concerned with such conceptual and epistemological questions. Instead, it seems that the specification of a Nash equilibrium is primarily driven by the need to find a solution to a game that is "in some sense, in equilibrium" (Varian, 1992, p. 264), or that satisfies a "natural consistency requirement" (p. 265). Hence, in an idealized setting a Nash equilibrium therefore corresponds to a "*steady state*" (Osborne, 2004, p. 20) in which there is no pressure for change. This, of course, is a familiar way of modeling economic phenomena for economists. Students of microeconomics typically learn about the theory of the market before learning about game theory: they study the concept of equilibrium by deriving it from the consumers' demand function and the firms' supply function, without conducting a thorough analysis of the epistemological

foundations of belief formation, and only *then* do they study the concept of a Nash equilibrium in the context of game theory and as a generalization of the Cournot equilibrium.[11] Thus, the derivation of a Nash equilibrium is a "natural" extension of how economists model aggregate outcomes as equilibria, or steady states.

This explanatory practice of economists distiguishes the notions of beliefs in the Nash equilibrium and in EUT. In the latter they are represented as a distribution of subjective probabilities over the state of the world, which follow certain rational requirements such as Bayesian updating and basic probability calculus. Subjective probability may even be weighted over objective probability, as in Tversky and Kahneman's Cumulative Prospect Theory (1992), because of the psychological principle of decreasing marginal sensitivity from certainty as a reference point, for example. EUT is a useful theory that allows psychologists to dovetail their insights because the concept of beliefs it espouses is a natural extension of the common-sense understanding with a formal representation, and psychologists are used to dealing with it. In contrast, beliefs in game theory—more specifically in a Nash equilibrium—do not correspond strongly with the intuitive understanding of beliefs: they have been derived from a discipline-specific drive to identify the properties of a system in (and out of) a steady state.

This drive is at the heart of the assumption that aggregate-level interactions among purposeful agents will, in the idealized condition, settle in some steady state. As a result, the implied notion of beliefs is not easily commensurable with the interpretations of psychologists, according to which they are either personal priors (the subjectivist view) or are based on some subject-independent features of the external world (the frequentist view). This explains why psychologists are typically unwilling to accept the legitimacy, let alone the usefulness, of a concept of beliefs that is so different from the concept they commonly adopt.

We do not mean to imply that game theorists are conceptually sloppy about what they mean by beliefs when we refer to the concept as ambiguous. On the contrary, beliefs are variably but precisely defined in Bayesian games, extensive games, and so on. The point is that the foundations of beliefs in game theory are primarily based on the practice of equilibrium analysis in economics in general, and the Nash equilibrium in game theory in particular. Within this paradigm, there are alternative ways of conceptualizing beliefs. For example, although the rationalizability approach "assumes that the players know each others' preferences, and considers what each player can deduce about the other players' actions from their rationality and their knowledge of each other's rationality" (Osborne, 2004, p. 21), Osborne clearly considers it optional and an alternative for economists. Economists, however,

[11] Although there is an alternative way of organizing microeconomics, proceeding from the optimizing individual to strategic interactions to market interactions (e.g. Bauman and Klein, 2010), such an organization seems still minority.

simply use a Nash equilibrium to model a steady state that will be reached through the interactions of experienced players. Ultimately, it is regarded as "a matter of judgment" (Osborne, 2004, p. 24) whether or not the notion is appropriate to model a given situation, not as a matter of whether or not epistemic foundations can be found, or the implied concept of beliefs can be reconciled with other existing belief concepts such as the subjectivist or the frequentist.

5 Game theory: a computational template as a boundary object

In this section, we apply *a computational template* to analyze the asymmetries between the ways EUT and GT functioned to facilitate interdisciplinary collaboration between economists and psychologists. Our observation that focusing on how TGEB as an innovative monograph was "diffused" across the disciplines (Doehne and Herfeld, 2018) cannot capture the asymmetries in which we are interested motivated our choice. First of all, we need a finer-grained unit of analysis than a publication of TGEB that includes both EUT and GT.

Second, our interest is not in the way in which TGEB was transferred from one domain to another, but in the conditions that allowed a community of researchers, including economists and psychologists, to work together and contribute to the shaping of the field of behavioral decision theory, and conversely in the factors that hindered the same research community to collaborate similarly to develop behavioral game theory. For these reasons, we think it is more fruitful to focus on a smaller unit of analysis (the computational template) and its function of *mediating* interdisciplinary collaboration rather than treating it as an object that was developed in one field and then was transferred to other domains. Let us start with the notion of templates.

Paul Humphreys (2004) was the first to use the notion of templates—in contrast to a theory and a model—as a unit of analysis to study progress in the physical sciences and their real-world applications. He notes that much progress in these fields is driven by the invention and deployment of tractable mathematical formulae, which he calls *computational templates*. A computational template is to be distinguished from a *theoretical* template in that it is a representational device limited to a mathematical or computational interpretation, whereas the theoretical template is interpreted within the domain of a theory (Humphreys 2018).

In addition to being mathematically tractable, computational templates "can be considered from a purely syntactic perspective" (Humphreys, 2004, p. 59), allowing some "flexibility and degree of independence from subject matter" (p. 67). Knuuttila and Loettgers (2014) identify this syntactic nature of computational templates as a

facilitator of their interdisciplinary transfer. Grüne-Yanoff (2011) similarly analyzes evolutionary game theory as an evolving template traveling from economics to biology, and back. Within this literature, interdisciplinary exchange is analyzed in terms of traveling computational templates, facilitated by their syntactic, subject-neutral nature.

We follow this line of thinking in focusing on computational templates as a unit of analysis. However, rather than seeing them as traveling syntactic objects, we find it more relevant here to construe them as objects that *mediate* coordination and collaboration between different scientific communities despite existing epistemic and conceptual tensions. This notion of "mediating" units goes back to the idea of *boundary objects* in the sociology of science, introduced by Star and Griesemer (1989) and further developed by researchers such as Wenger (1998). According to Star and Griesemer (1989):

> "Boundary objects are objects which are both plastic enough to adapt to local needs and the constraints of the several parties employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use." (p. 393)

What is curious about GT and EUT is that prima facie they share the features that would make them both boundary objects and computational templates: given that the variables they adopt, i.e., beliefs, preferences, and choice, are the same, they should be equally applicable across the same domains. At the very least, if success is achieved in the case of EUT and psychology, the same should apply to the case of GT and psychology. Moreover, with respect to the formalism they employ, they also seem to share a similar mathematical machinery, which would suggest that they should be equally suitable as means of interdisciplinary transfer.

However, we maintain that EUT works well as a boundary object and a computational template in economics and psychology, but that the same does not apply to GT. Our explanation of asymmetric collaboration can be summarized thus: as a boundary object, EUT facilitates collaboration between economists and psychologists because the *interpretation* of the psychological concepts employed in EUT (beliefs, preferences and choices) is largely shared between these two communities (as well as among philosophers and lay people). They differ in terms of practices such as experimental procedures, but nevertheless, psychologists were heavily involved in modifying EUT in the face of anomalous empirical findings. The level of abstraction was not an obstacle, but rather presented an opportunity for psychologists to contribute their expertise in modifying the model. Psychologists probably have less faith in individual optimization than economists have, but the ideal type of optimizing individuals was abstract enough to allow their involvement.[12]

[12]Of course, some psychologists refuse to collaborate with economists precisely because they

Rather than focusing on different styles of modeling (economists' optimization-based vs. psychologists' process-based modeling), we highlight the centrality of equilibrium analysis as economic practice, and the implied notion of beliefs in equilibrium, which constitutes a conceptual obstacle hindering psychologists from participating in the modification of GT drawing on psychological expertise.

In this sense, EUT functioned as a boundary object that was flexible enough to be used by economists and psychologists with different theoretical and experimental backgrounds, while at the same time its substantive interpretation was robustly shared to allow for fruitful collaboration. On the other hand, the requirements on GT regarding use of the concept of beliefs, which derive from the formal requirements of equilibrium-based analysis, have no counterpart in psychology. Although GT as a computational template traveled across many domains, it did not function as a fruitful boundary object between psychology and economics: the belief concept was not robust enough to allow for overlapping interpretations.

6 Conclusions

The interdisciplinary collaboration between psychologists and economists mediated by EUT gave rise to behavioral decision theory, which in turn has formed a core part of behavioral economics. The empirical and theoretical work of cognitive psychologists was crucial in this development(Heukelom, 2014). The exchange took place as a modification of EUT as a boundary object. Game theory did not function in the same way, despite its common origin. In particular, the extent to which insights from psychology have modified GT is limited, although GT did travel to psychology in the form of experimental paradigms such as the Prisoner's Dilemma and other social dilemmas. As a consequence, behavioral game theory as a subfield of behavioral economics remains, for the most part, a game for economists.

We have explained this asymmetry as a result of the peculiarities of equilibrium analysis in the practice of economists, which embed the conceptual differences in how economists and psychologists understand beliefs—a seemingly unproblematic notion. If our explanation is on the right track, there are implications for the methodology of interdisciplinary sciences. Specifically, those answering a common call for a more interdisciplinary approach in the behavioral and social sciences—which we endorse and encourage—will need to pay more careful attention to the conceptual differences that may constitute a barrier to this admirable cause. Gintis (2017) identifies the reluctance of non-economists to fully embrace common-core analytical foundations—such as decision theory and game theory—as the main obstacles against integrating or unifying the behavioral sciences (meaning the social sciences plus sociobiology).

reject optimization as even an ideal type. Our point is that EUT was abstract enough that critical mass of psychologists collaborated.

However, this reluctance may not be solely attributable to disciplinary rent-seeking or a lack of mathematical training on the part of psychologists and other social scientists: conceptual differences may play a key role as a specific kind of what MacLeod (2016) refers to as *cognitive obstacles* to interdisciplinarity. Recognizing such a conceptual difference is a first step in designing more effective interdisciplinary methodological strategies.

References

Bacharach, M., Gold, N., and Sugden, R. (2006). *Beyond individual choice: teams and frames in game theory*. Princeton University Press, Princeton, N.J.

Bauman, Y. and Klein, G. (2010). *The Cartoon Introduction to Economics: Volume One: Microeconomics*. Hill and Wang.

Camerer, C. F. (2003). *Behavioral Game Theory*. Princeton University Press, Princeton, NJ.

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences*, 26:139–+.

Dawes, R. M., McTavish, J., and Shaklee, H. (1977). Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology*, 35(1):1–11.

Dhami, S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford University Press, Oxford.

Doehne, M. and Herfeld, C. (2018). The diffusion of scientific innovations: A role typology. *Studies in History and Philosophy of Science Part A*.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3):817–868.

Gintis, H. (2017). *Individuality and Entanglement: The moral and material bases of social life*. Princeton University Press, Princeton, NJ.

Giocoli, N. (2003). *Modeling rational agents: From interwar economics to early modern game theory*. Edward Elgar Publishing, Cheltenham, UK.

Grether, D. and Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review*, 69:623–638.

Grüne-Yanoff, T. (2011). Models as Products of Interdisciplinary Exchange: Evidence From Evolutionary Game Theory. *Studies in History and Philosophy of Science Part A*, 42(2):386–397.

Grüne-Yanoff, T. and Lehtinen, A. (2012). Philosophy of game theory. *Philosophy of economics*, pages 531–576.

Guala, F. (2006). Has game theory been refuted? *The Journal of Philosophy*, 103(5):pp. 239–263.

Hands, D. W. (2010). Stabilizing consumer choice: the role of 'true dynamic stability' and related concepts in the history of consumer choice theory. *The European Journal of the History of Economic Thought*, 17(2):313–343.

Heukelom, F. (2014). *Behavioral economics: a history*. Cambridge University Press, Cambridge.

Humphreys, P. (2004). *Extending ourselves: Computational science, empiricism, and scientific method*. Oxford University Press.

Isaac, R. M., McCue, K. F., and Plott, C. R. (1985). Public goods provision in an experimental environment. *Journal of Public Economics*, 26(1):51 – 74.

Kadane, J. B. and Larkey, P. D. (1982). Subjective probability and the theory of games. *Management Science*, 28(2):113–120.

Kahneman, D. (1988). Experimental economics: A psychological perspective. In Tietz, R., Albers, W., and Selten, R., editors, *Bounded Rational Behavior in Experimental Games and Markets*, pages 11–20, Berlin. Springer-Verlag.

Kahneman, D. (2014). Daniel Kahneman - Biographical. *Nobel Media AB*.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986). Fairness and the assumptions of economics. *The Journal of Business*, 59(4):pp. S285–S300.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of political Economy*, 98(6):1325–1348.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):pp. 263–292.

Kim, O. and Walker, M. (1984). The free rider problem: Experimental evidence. *Public Choice*, 43(1):3–24.

Knuuttila, T. and Loettgers, A. (2014). Magnets, spins, and neurons: The dissemination of model templates across disciplines. *The Monist*, 97(3):280–300.

Ledyard, J. (1995). Public goods: a survey of experimental research. In Kagel, J. H. and Roth, A. E., editors, *The handbook of experimental economics*, pages 111–194. Princeton University Press.

Lichtenstein, S. and Slovic, P. (1971). Reversals of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology*, 89:46–55.

Lisciandra, C. (2018). The role of psychology in behavioral economics: The case of social preferences. *Studies in History and Philosophy of Science Part A*, 72:11 – 21.

MacLeod, M. (2016). What makes interdisciplinarity difficult? some consequences of domain specificity in interdisciplinary practice. *Synthese*.

Mnookin, R. H., Ross, L., Arrow, K. J., and Tversky, A., editors (1995). *Barriers to conflict resolution*. Norton New York, NY.

Morris, S. (1995). The common prior assumption in economic theory. *Economics and Philosophy*, 11(2):227–253.

Nagatsu, M. and Małecka, M. (2019). How behavioural research has informed consumer law: the many faces of behavioural research. In Micklitz, H.-W., Sibony, A.-L., and Esposito, F., editors, *Research Methods in Consumer Law: A Handbook*, Handbooks of Research Methods in Law series, chapter 11, pages 357–398. Edward Elgar Publishing, Cheltenham, UK.

Osborne, M. J. (2004). *An introduction to game theory*. Oxford University Press, Oxford.

Star, S. L. and Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in berkeley's museum of vertebrate zoology, 1907-39. *Social Studies of Science*, 19(3):387–420.

Tversky, A. and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2):105–110.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of utility. *Journal of Risk and Uncertainty*, 5:297–323.

Varian, H. R. (1992). *Microeconomic Analysis*. W.W. Norton & Co., 3rd edition.

Varian, H. R. (2010). *Intermediate Microeconomics; a modern approach*. 8th edition.

Von Neumann, J. and Morgenstern, O. (2004). *Theory of games and economic behavior*. Princeton University Press, Princeton, N.J., 60th anniversary edition.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press, Cambridge, UK.